Assessing Material Aging from Doubly Censored Data : Weibull Distribution vs. Poisson Process

Gilles Celeux, Christian Lavergne, Yann Vernaz

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Assessing material aging from doubly censored data: Weibull distribution vs. Poisson process

Gilles Celeux, Christian Lavergne, Yann Vernaz

N° 3857

Janvier 2000

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--3857--FR+ENG

Assessing material aging from doubly censored data: Weibull distribution vs. Poisson process

Gilles Celeux, Christian Lavergne, Yann Vernaz

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IS2

Rapport de recherche n° 3857 — Janvier 2000 — 21 pages

Abstract: The versatile Weibull distribution is popular for modeling aging in failure time problems. However, in some situations the only available data are right or left censored data and estimating the Weibull distribution parameters is made much more difficult. In this paper, we consider the performance of estimating a Weibull distribution in such a doubly censored context from the maximum likelihood approach and from a non informative Bayesian point of view. Moreover, we propose an alternative model for assessing aging. It consists in assuming that left censored data arise from a Poisson process. This model can appear to provide more reliable results from such poorly informative failure time data. Both approaches are compared on the basis of numerical experiments.

Key-words: Failure time data, right and left censored data, Weibull distribution, maximum likelihood and Bayesian estimates, Poisson process, aging effect.

(Résumé : tsvp)

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Modélisation du vieillissement d'un matériel à partir de données doublement censurées: loi de Weibull vs. processus de Poisson

Résumé : On s'intéresse à la modélisation de la durée de vie d'un matériel réparable à partir de données doublement censurées. Aucune date de défaillance n'est observée. La loi de Weibull est un modèle couramment employé pour modéliser la durée vie d'un matériel. Cependant, eu égard à la nature des données, l'estimation des paramètres de la distribution de Weibull s'avère difficile voire impraticable. Nous étudions la performance des estimations dans le contexte d'une inférence directe par la méthode du maximum de vraisemblance puis par une approche bayésienne non informative. Les résultats numériques obtenus sont peu satisfaisants et nous amènent à proposer une alternative à la modélisation de Weibull. Notre méthode consiste à traiter une défaillance comme un événement rare et à modéliser le vieillissement par palier suivant un processus de Poisson. On associe ensuite le processus de Poisson avec un modèle GLM (Generalized Linear Model). Cette méthodologie se révèle mieux adaptée dans ce cadre, et permet de donner des réponses simples et fiables aux questions que l'on se pose, à savoir, déceler un vieillissement ou la présence d'un effet de conception du matériel.

Mots-clé : Temps de défaillance, données censurées à droite et à gauche, loi de Weibull, maximum de vraisemblance, bayésien non informatif, processus de Poisson.

1 Introduction

The Weibull distribution provides a good and largely employed model for the life of electrical and mechanical systems among others. In particular, the Weibull distribution is well suited for taking into account the senescence of a material.

However, in many circumstances the shape and scale parameters of the Weibull distribution must be estimated from highly right censored samples providing very few failure times. Such samples jeopardized reliable inference from this distribution. In some situations, things are worse: there is no observed failure times and the only available data are right or left censored data, and consequently Weibull inference is made much more difficult. This case is not unusual. It happens for instance in a circumstance that we consider in this article:

- A material is composed of numerous identical and independent components whose failure does not imply the failure of the whole system but a damaging running mode.
- The controls of the materials are achieved at fixed periodic times. At each control, it is observed if some of the components have failed since the previous control. But the exact failure times are unknown.
- At each control, the material is maintained and it is considered *as good as new* after each control.

The seminal idea of this paper is that it is difficult to derive reliable estimates of the Weibull distribution parameters in such a context and there is the need for alternative modeling even if the Weibull distribution is the correct model underlying the damaging process. From this point of view, it seems natural to deal with a Poisson process in this doubly censored setting. This model consists of assuming that the left censored data between two control times are the realizations of a Poisson distribution. This choice is supported by the fact that material damaging is a rare event. In this context, aging can be expressed by assuming that the Poisson process is inhomogeneous, namely that the parameter of the Poisson distribution depends on the period of control. Thus, aging effect is assessed in a simple and reliable way via a generalized linear model derived from the Poisson process.

This article is devoted to the comparison of both models, Weibull distribution and inhomogeneous Poisson process, for assessing aging when the only available data are left or right censored data. It is organized as follows. Data-type we are concerned with is described Section 2. Maximum likelihood and non informative Bayesian inference for the parameters of a Weibull distribution in this context are presented in Section 3. The modeling of aging with a generalized linear model derived from a Poisson process is presented in Section 4. Section 5 is devoted to numerical experiments comparing both approaches and a discussion section ends the article.

2 Data and notation

The lifetime data that we considered are as follows. N identical materials are available, and each material is composed of M identical and independent units working in series. We are concerned with times to failure of the units (or components) of a material. It is important to notice that the failure of a component does not imply the failure of the material. It just causes a damaging running mode of the material. Consequently the exact times to failure of the units are not available. Moreover, it is assumed that at each control all failed components are replaced. Thus after each control, materials are considered *as good as new.*

Controls are realized at fixed times for each component $j = 1, ..., M$ of material $i = 1, \ldots, N$. The period for the kth control for material i is denoted t_i^k for $i = 1, \ldots, N$ and $k = 1, \ldots, K$, $K \geq 2$.

Remark: Since materials are assumed to be *as good as new*, assessing aging can be done if the t_i^k's are not equal. In the following we assume that the period control lengths t_i^k, $k = 1, ..., K$ are different for a same material i. If materials were *as bad as old*, the t_i^k could be equal without inconvenient and a similar analysis could be performed. The only difference will consists in the likelihood expressions for both Weibull modeling and Poisson modeling would be different. For brevity, we do not consider this case in this article.

At control k, $1 \leq k \leq K$, the number of left censored data for material i is denoted m_i^k and $m^k = \sum_{i=1}^n m_i^k$ denotes the total number of left censored data. Thus $M - m_i^k$ is the number of right censored data for material i and $N \times M - m^k$ is the total number of right censored data at control k. Denoting $t_{i,j}^k$ the discovery date of failure for component $j = 1, ..., M$ of material $i = 1, ..., N$, data obtained at control k can be represented in the following way:

$$\mathcal{D}^k = (\underbrace{\underbrace{t_{1,1}^k, ..., t_{1,m_1^k}^k}_{\text{left cens. da.}}, \underbrace{t_{1,m_1^k+1}^k, ..., t_{1,M}^k}_{\text{right cens. da.}}}_{\text{units of material 1}}, ..., \underbrace{\underbrace{t_{N,1}^k, ..., t_{N,m_N^k}^k}_{\text{left cens. da.}}, \underbrace{t_{N,m_N^k+1}^k, ..., t_{N,M}^k}_{\text{right cens. da.}}}_{\text{units of material } N}).$$

Since we have for all $j = 1, \ldots, M$, $t_{i,j}^k = t_i^k$ for all the components of material i, the data $\mathcal{D}^k$ consists of m^k left censored data:

$$\underbrace{t_1^k, \ldots, t_1^k}_{m_1^k \text{ times}}, \ldots, \underbrace{t_N^k, \ldots, t_N^k}_{m_N^k \text{ times}}$$

and $N \times M - m^k$ right censored data:

$$\underbrace{t_1^k, \ldots, t_1^k}_{M-m_1^k \text{ times}}, \ldots, \underbrace{t_N^k, \ldots, t_N^k}_{M-m_N^k \text{ times}}.$$

3 Weibull modeling

The Weibull distribution $\mathcal{W}(\beta, \eta)$ with shape parameter β and scale parameter η is much used for analyzing times to failure data (see for instance Lawless 1982). Denoting $\theta = (\beta, \eta)$, its probability density function is:

$$f(t/\theta) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \text{ for } t \geq 0,$$

and its survivor function is:

$$R(t/\theta) = \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \quad \text{for } t \geq 0.$$

In the Weibull modeling, a shape parameter $\beta > 1$ produces an increasing hazard function and is evidence for aging. Thus a good estimate of β is crucial for assessing aging in a reliable way from the Weibull distribution.

3.1 Direct likelihood inference

In this context the likelihood function of the parameter $\theta = (\beta, \eta)$ for data $\mathcal{D} = (\mathcal{D}^k, k = 1, \ldots, K)$ takes the form:

$$L(\mathcal{D}; \theta) = \prod_{k=1}^{K}\prod_{i=1}^{N}(1 - R(t_i^k/\theta))^{m_i^k} R(t_i^k/\theta)^{M-m_i^k}.$$

The log-likelihood function is written

$$\log L(\mathcal{D}; \theta) = \sum_{k=1}^{K}\sum_{i=1}^{N} m_i^k \log\left[1 - \exp[-(\frac{t_i^k}{\eta})^{\beta}]\right] - \sum_{k=1}^{K}\sum_{i=1}^{N}(M - m_i^k)\left(\frac{t_i^k}{\eta}\right)^{\beta}. \tag{1}$$

The maximum likelihood (ml) estimator is found by maximizing equation (3.1) with respect of parameter θ. Then, it is derived by solving the following system of equations (see Appendix A).

$$\sum_{k=1}^{K}\sum_{i=1}^{N}(t_i^k)^{\beta}\left[(m_i^k - M) - m_i^k \frac{R(t_i^k/\theta)}{1 - R(t_i^k/\theta)}\right] = 0 \tag{2}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{N}(t_i^k)^{\beta}\log(t_i^k)\left[(m_i^k - M) - m_i^k \frac{R(t_i^k/\theta)}{1 - R(t_i^k/\theta)}\right] = 0 \tag{3}$$

Those likelihood equations are highly non-linear in θ, the solution is achieved by numerical techniques. In particular, those equations can by solved by using an iterative procedure as the Newton-Raphson algorithm. Unfortunately, this method can fail due to the nature of the doubly censored data. And, numerical trouble is increased by the fact that the censoring times are the same.

Typically, we are facing a missing data structure model, the EM algorithm of Dempster, Laird and Rubin (1977) can be thought of as useful to derive the ml estimator. In fact, we proved in Appendix B that, the fixed points of EM are solutions of the likelihood equations (2) and (3) and consequently the EM algorithm is of no help in the present context.

A possible way to bypass this numerical limitation is to replace left censored data with fictitious failure times. Usual strategies of filling left censored data are the following (see for instance Lannoy et Procaccia 1994).

1. Consider left censored times as failure times.
2. Replace left censored times by some predefined time as for instance the medium of interval of times between $(t_0 + t'_0)/2$ and t_0 (t_0: current control time, t'_0: previous control time).
3. Replace left censored times by a random time drawn from a uniform distribution on $[(t_0 + t'_0)/2, t_0]$.

In this context, denoting $\tilde{t}^k_{i,j}$ the fictitious failure times, the likelihood takes the form:

$$L(\mathcal{D};\theta) = \prod_{k=1}^{K}\prod_{i=1}^{N}\left[\prod_{j=1}^{m_i^k} f(\tilde{t}^k_{i,j}) \prod_{j=m_i^k+1}^{M} R(t^k_{i,j})\right].$$

and leads to the classical Weibull likelihood equations for right censored data (see for instance Lawless 1982). Those equations are easier to solve than equations (2) and (3) even in a small sample setting and do not involve the possible numerical trouble encountered with (2) and (3) as it will be illustrated in the numerical experiments.

3.2 Non informative Bayesian inference

An alternative to maximum likelihood estimation is to adopt a Bayesian approach. In this approach the parameter θ is regarded as random with prior distribution $\pi(\theta)$ and the inference on θ is based on the posterior distribution

$$\pi(\theta|\mathcal{D}) = \frac{\pi(\theta)L(\mathcal{D};\theta)}{\int \pi(\theta)L(\mathcal{D};\theta)d\theta}. \tag{4}$$

In our context no prior information is available, we simply view the Bayesian paradigm as a way to get regularized estimates. For this reason, the Bayesian estimate that we consider is the mean of the posterior distribution $\pi(\theta|\mathcal{D})$ which is optimal for a quadratic loss function (see for instance Robert 1994) and we choose a non informative prior distribution. Choosing a non informative prior distribution for a multivariate parameter is nothing but obvious. A recommendable reference for a discussion on the choice of non informative priors for two-parameter Weibull distributions is Sun (1997). In this paper, Sun considers the Jeffreys prior which is $\pi(\theta) = 1/\eta$ and the reference prior $\pi(\theta) = 1/(\beta\eta)$. Sun advocates the use of the reference prior rather than the Jeffreys prior.

In our doubly censored context, for any prior distribution the posterior mean has to be approximated through numerical integration. In our experiments we make use of both non informative prior distributions and we approximate the posterior mean of θ through Monte Carlo integration.

4 Poisson Modeling

Data presented in Section 2 can be regarded as realizations of a point process. We propose to model the number of failed components as a Poisson process and to assess aging from inference on the intensity of the Poisson process.

Let $\{C_t, t \geq 0\}$ denote the counting process of the number of damaged materials before t with $C_0 = 0$. Assuming no aging, random variable $S_t = C_{t+s} - C_s$ is independent of s and is unaffected by what happens prior to time t. The distribution of S_t is a binomial distribution $\mathcal{B}(N \times M, p(t) = \mu t)$. Since typically the number of damaged materials is quite small, the classical approximation of the binomial distribution by a Poisson distribution is valid (see for instance Feller 1968 pp.153) and we assume that

$$P\left(S_t = m\right) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}.$$

where the parameter $\lambda = MN\mu$ of the Poisson distribution $\mathcal{P}(\lambda t)$ does not depend of time t. It means that $\{S_t, t \geq 0\})$ is a homogeneous Poisson process with intensity λ not depending of time. In addition, times between occurrences of damaging are i.i.d. exponential random variables with failure rate λ. For this model the likelihood of λ with data $\mathcal{D}$ is

$$\mathcal{L}(\mathcal{D}; \lambda) = \prod_{k=1}^{K} \prod_{i=1}^{N} \frac{(\lambda t_i^k)^{m_i^k}}{m_i^k!} e^{-\lambda t_i^k},$$

From which it easily follows that the m.l. estimate of λ is

$$\hat{\lambda} = \frac{\sum_{k=1}^{K} \sum_{i=1}^{N} m_i^k}{\sum_{k=1}^{K} \sum_{i=1}^{N} t_i^k}.$$

On the contrary, assuming aging means that the intensity λ of the inhomogeneous Poisson process is time-dependent. More precisely, we assume that the total number of failed components at control k ($m^k = \sum_{i=1}^{N} m_i^k$) is the realization of a Poisson distribution with parameter λ_k depending of the period of control. Namely, the events observed at control k

are realizations of a homogeneous Poisson process with intensity λ_k depending of the period k, i.e. $m_i^k \sim \mathcal{P}\left(\lambda_k t_i^k)\right)$. For this model, the likelihood of λ_k with data $\mathcal{D}^k$ is

$$\mathcal{L}(\mathcal{D}^k;\lambda_k) = \prod_{i=1}^{N} \frac{(\lambda t_i^k)^{m_i^k}}{m_i^k!} e^{-\lambda t_i^k},$$

from which it follows that

$$\hat{\lambda}_k = \frac{\sum_{i=1}^N m_i^k}{\sum_{i=1}^N t_i^k}.$$

Aging as a fixed effect in a Generalized Linear Model For assessing absence of aging from this modeling, we can consider the problem in the framework of generalized linear model. In the Poisson modeling, aging can be expressed as a fixed effect of the period control length with an anova-like equation

$$\lambda_k = \lambda + \alpha_k \tag{5}$$

where λ is a general mean effect and α_k represents the effect of the period of control length (*aging* effect). This, model can be dealt with using the Generalized Linear Model (GLM) theory of McCullagh and Nedler (1989). The essential difference with classical analysis of variance is that the observed variables are distributed according to the Poisson distribution rather than the Gaussian distribution (see McCullagh and Nedler, 1989 p. 30). In this GLM framework we consider the hypothesis test H_0 ($\alpha^k = 0$ for all k) against H_1 (there exists at least one k such that $\alpha^k \neq 0$). A likelihood ratio test is designed. The test statistic is

$$\xi = 2\left[\log \mathcal{L}(\mathcal{D};\hat{\lambda}_k) - \log \mathcal{L}(\mathcal{D};\hat{\lambda}_0)\right]$$

where $\mathcal{L}(\mathcal{D};\hat{\lambda}_k)$ (resp. $\mathcal{L}(\mathcal{D};\hat{\lambda}_0)$) is the likelihood under H_1 (resp. H_0). Under classical regularity assumptions McCullagh and Nedler 1989), the test with asymptotic level α is defined by the critical region: $\left\{\xi \geq \chi^2_{1-\alpha}(K-1)\right\}$, where $\chi^2_{1-\alpha}(K-1)$ is the quantile $1-\alpha$ of the χ^2 distribution with $K-1$ degrees of freedom (df).

An other great interest to embed the problem in a generalized linear model framework is that it allows to assess other fixed effect in the same exercise. For instance suppose the N materials are coming from different manufacturers. Then we can asses both the aging effect and the manufacturer effect. This is done with the equation

$$\lambda_{k\ell} = \lambda + \alpha_k + \beta_\ell + \gamma_{k\ell}$$

where λ is the general mean effect, α_k is the aging effect, β_ℓ is the fixed effect of the material manufacturer, and γ_{jk} is the interaction between the aging effect and the manufacturer effect. Again, this model can be analyzed with GLM tools. A numerical example where a *material* effect is considered in addition to the aging effect is presented in the next section. In the following, GLM arising from the Poisson process will be denoted Poisson-GLM.

5 Numerical experiments

In this section, we present Monte Carlo simulations to test the ability of the above mentioned methods for assessing aging when the only available data are right or left censored data. In the first Monte Carlo numerical experiment, we consider that aging is controlled with a Weibull distribution. In a second Monte Carlo experiment aging is controlled from an alternative distribution. Finally, an additional example is presented to illustrate the Poisson-GLM ability to detect an alternative fixed effect to aging.

5.1 Weibull experiments

We consider failure times arising from a Weibull distribution. Each considered situation was replicated 100 times. We consider $N = 10$ materials and each material is composed of $M = 100$ identical and independent units. We consider two different control times $c_1 = 100$ and $c_2 = 200$. Thus the total sample size is $2 \times N \times M = 2000$. Convergence of the iterative algorithms is appreciated for an accuracy of 10^{-6} with a maximum of 5000 iterations. Beyond that, we consider that there is not convergence.

We have considered four different Weibull distribution $\mathcal{W}(\beta, \eta)$ with $(\beta = 1, \eta = 4000), (\beta = 1.5, \eta = 1500), (\beta = 2, \eta = 800)$, and $(\beta = 3, \eta = 500)\}$. Table 1 provides the probability of failure before the censoring times c_1 and c_2 for the four simulated models. First we

Table 1: Probability of failure before censoring times c_1 and c_2.

censoring time		$c_1 = 100$	$c_2 = 200$
Model	$\mathcal{W}(1, 4000)$	0.0247	0.0488
Model 2	$\mathcal{W}(1.5, 1500)$	0.0171	0.0475
Model 3	$\mathcal{W}(2, 800)$	0.0155	0.0606
Model 4	$\mathcal{W}(3, 500)$	0.0080	0.0620

displayed in Table 2 the results obtained using the three left censored replacing strategies presented in Section 3.1. Clearly, those strategies provide quite biased estimates and cannot be recommended.

Table 3 displays the results for the direct ml estimator, the non informative Bayesian estimator and the aging test from the Poisson-GLM model. For the ml and Bayes estimators

Table 2: ml estimates from the replacing strategies

	mle S1 mean (std)	mle S2 mean (Std)	mle S3 mean (std)
$\beta = 1$	4.96 (0.71)	2.43 (0.16)	1.03 (0.11)
$\eta = 4000$	346.6 (28.0)	632.1 (65.5)	4151.7 (1627.3)
$\beta = 1.5$	6.00 (1.22)	2.63 (0.20)	1.07 (0.16)
$\eta = 1500$	324.7 (29.7)	606.0 (61.5)	4461.4 (2447.2)
$\beta = 2$	8.07 (2.16)	2.90 (0.20)	1.12 (0.14)
$\eta = 800$	281.9 (22.9)	506.9 (45.4)	3060.2 (1378.5)
$\beta = 3$	15.75 (8.90)	3.30 (0.20)	1.20 (0.16)
$\eta = 500$	244.64 (16.8)	458.2 (28.7)	2811.8 (1449.4)

we give the mean value and into parentheses the standard error over the 100 simulations. For the Poisson-GLM test of aging, we give the percentage of times aging is accepted at a standard 5% level over the 100 simulations. Before commenting this table, it is important to notice that the ml method appears to be quite sensitive to the doubly censored context: the Newton-Raphson algorithm did not converge in 57% of cases for model $\mathcal{W}(1, 4000)$, 73% for $\mathcal{W}(1.5, 1500)$, 71% for $\mathcal{W}(2, 800)$ and 74% for $\mathcal{W}(3, 500)$. (Using other optimization algorithms as conjugate gradient did not lead to better results.) On the contrary, both

Table 3: Estimates using ml, Bayes inference and GLM aging test

	ml mean (std)	Bayes mean (std)	GLM percentage of accepting aging
$\beta = 1$	1.12 (0.33)	1.28 (0.18)	7%
$\eta = 4000$	11114.6 (32768.9)	2546.7 (574.4)	
$\beta = 1.5$	1.49 (0.40)	1.58 (0.32)	24%
$\eta = 1500$	2859.2 (5913.5)	1932.0 (653.6)	
$\beta = 2$	2.12 (0.45)	2.09 (0.47)	84%
$\eta = 800$	839.8 (343.3)	1021.0 (431.5)	
$\beta = 3$	3.20 (0.71)	3.16 (0.70)	100%
$\eta = 500$	499.3 (102.0)	548.0 (141.6)	

the non informative Bayesian approach and the Poisson-GLM modeling do the job. When converging, the ml estimation is reasonable even if the scale parameter η can be dramatically overestimated when the shape parameter β is near 1. But the Bayesian methodology can be preferred since it provides a more precise insight on the lifetime distribution. However,

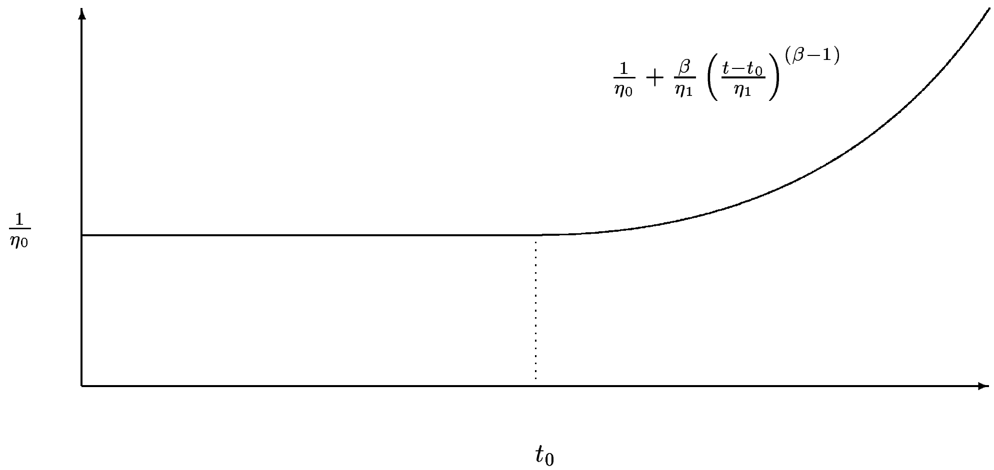

Figure 1: The hazard function of the alternative model

it is important to notice that this first experiment favors modeling aging with Weibull distribution. In the second experiment, we analyze data for which aging is not associated to a Weibull distribution.

5.2 An alternative model

We consider an alternative aging model defined with the following hazard function

$$h(t) = \frac{1}{\eta_0} + \frac{\beta}{\eta_1}\left(\frac{x-t_0}{\eta_1}\right)^{(\beta-1)} \mathbf{I}_{\{t>t_0\}}(t),$$

where $\mathbf{I}_A(.)$ denotes the indicator function of the set A. Its hasard function is depicted in Figure 1. This model is easily simulated since it can be shown that a lifetime X drawn according this model can be written as $X = \min(Y, W)$ where Y is an exponential random variable with scale parameter η_0 and W is a three parameter Weibull distribution with shape parameter β, scale parameter η_1 and position parameter t_0. In what follows, we denote $\mathcal{B}e(t_0, \theta)$ with $\theta = (\beta, \eta_0, \eta_1)$ this distribution.

We simulated data sets from this distribution with $t_0 = 100$, $\theta = (1.5, 800, 1500)$. As previously, we consider $N = 10$ materials and each material is composed of $M = 100$ identical and independent units. We consider two different control times $c_1 = 100$ and $c_2 = 200$. The total sample size is $2 \times N \times M = 2000$ and each considered situation was replicated 100 times.

The results are displayed in Table 4. Recall that the ml and Bayes estimators are derived assuming a Weibull assumption, not the true $\mathcal{B}e$ distribution. First, notice that the ml procedure failed to converges in 57% of replications. The ml and Bayesian estimators are clearly unable to detect aging. Moreover the ml estimate of the scale parameter η is quite unstable. On the contrary, The Poisson-GLM modeling detects an aging effect in a majority of replications.

Table 4: Estimates using ml, Bayes inference and GLM aging test for the $\mathcal{B}e$ distribution

	ml mean (std)	Bayes mean (std)	GLM percentage of accepting aging
β	0.96 (0.28)	1.04 (0.15)	60%
η	10611.9 (55494.8)	2300.1 (625.3)	

5.3 An example with an alternative effect

We consider a simple example with no aging effect but with a "material" effect. Again We consider $N = 10$ materials and each material is is composed of $M = 100$ identical and independent units. We consider two different control times $c_1 = 100$ and $c_2 = 200$. But five materials have a $\mathcal{W}(1, 2000)$ distribution and five materials have a $\mathcal{W}(1, 4000)$ distribution. The resulting data set is displayed in Table 5.

Table 5: Five materials from $\mathcal{W}(1, 4000)$ and five materials from $\mathcal{W}(1, 2000)$.

Censoring time	$\mathcal{W}(1, 4000)$ Left censored components	$\mathcal{W}(1, 2000)$ Left censored components
100	1	6
100	3	8
100	1	5
100	4	5
100	5	6
200	5	9
200	5	10
200	3	7
200	7	13
200	3	8

Table 6 provides ml and Bayesian Weibull parameter estimates for both groups of materials. None of the methods is able to detect that the material groups are essentially different.

Table 6: ml and Bayesian Weibull parameter estimates for both groups of materials.

Method	Model	Parameter	Estimation
ml	$\mathcal{W}(1, 4000)$	β	0.7296
		η	6179.5
	$\mathcal{W}(1, 2000)$	β	0.6739
		η	6211.8
Bayesian	$\mathcal{W}(1, 4000)$	β	1.2198
		η	2657.9
	$\mathcal{W}(1, 2000)$	β	0.9697
		η	2501.2

Table 7: GLM results for the two group data set.

Model	Parameter	Estimation (std)	Deviance	df
Model 0	λ	3.8000 (0.3559)	26.24	19
Model 1	λ	4.4000 (0.6633)	24.85	18
	α	-0.9000 (0.7842)		
Model 2	λ	2.4667 (0.4055)	11.90	18
	β	2.6667 (0.7118)		
Model 3	λ	2.9970 (0.6721)	10.76	17
	α	-0.7702 (0.7379)		
	β	2.6329 (0.7088)		

Table 7 displays the results of the Poisson-GLM modeling. The following models were considered:

- Model 0: No fixed effect: λ.
- Model 1: *Length of period* effect: $\lambda + \alpha$.
- Model 2: *Material* effect: $\lambda + \beta$.
- Model 3: *Length of period* and *material* effects: $\lambda + \alpha + \beta$.

We evaluate the different effect using hypotheses test at a $\alpha = 5\%$ asymptotic level. Denoting D_i the deviance of model i, $i = 0, 1, 2, 3$, we have $D_0 - D_1 = 1.39$, $D_0 - D_2 = 14.34$ and $D_2 - D_3 = 1.14$. Those values have to be compared to the p-value $\chi^2_{1-\alpha}(1) = 3.84$. Consequently, we conclude positively for model 2 (material effect) which is the right model.

6 Discussion

Assessing aging exclusively from left and right censored data is a difficult task. In this article, we have considered this problem from different points of view. A classical way for assessing aging is to assume that the failure times arise from a Weibull distribution. In such a doubly censored context, our experiments show that non informative Bayesian inference can be preferred to the maximum likelihood methodology which can often encounter numerical troubles and provide less stable estimators. But the main point of our contribution is that is more natural and more sensible to view aging as a GLM fixed effect of a Poisson process. Numerical experiments show that this way of modeling aging appears to be more reliable and more secure.

We limit our numerical investigations to materials considered as good as new. In this setting, the lengths of control have to be different to make the analysis of aging possible. When materials can be considered *as bad as old*, there is not such a limitation and the Poisson-GLM modeling is also more natural.

However, it is important to keep in mind that assessing aging when only left and right censored data are available remains a poorly posed problem. For instance, we simulated according to the same experimental scheme of Section 5.3, five materials from a $\mathcal{W}(1, 4000)$ and five materials from a $\mathcal{W}(2, 2000)$. Since the repartition of left and right censored data was almost the same for the two Weibull distributions at censoring times $c_1 = 100$ and $c_2 = 200$, it is rather impossible to distinguish them and none of the above mentioned methods was able to detect both the aging and the material effect.

Acknowledgments We thank Marie-Agnès Garnero and Benjamin Villain for introducing us to this problem and for helpful discussions, and Henri Bertholon for its help when treating the alternative model of Section 5.2.

References

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal. Roy. Statist. Soc. (Ser. B)*, **39**, 1-38.

Feller, W. (1968) *An Introduction to Probability Theory and its Applications,* (second edition). Wiley, New York.

Lawless, J. F. (1982) *Statistical models and methods for lifetime data.* Wiley, New York.

Lannoy, A. and Procaccia H. (1994) *Méthodes avancées d'analyse des bases de données du retour d'expérience industriel.* Eyrolles, Paris.

McCullagh P. and Nelder J. A. (1989) *Generalized Linear Models,* (second edition). Chapman and Hall, London.

Sun, D., (1997) A note on non informative priors for Weibull distributions. *Journal of Statistical Planning and Inference*, **61**, 319-338.

Robert, C. P., (1994) *The Bayesian Choice.* Springer-Verlag, New York.

Appendix A: Likelihood Equations

For simplicity, we omit the index k and consider a unique period of control. Generalization is straightforward. The likelihood function with $N \times m_i$ left censored times t_i and $N \times (M - t_i)$ right censored times t_i is

$$\log L(\mathcal{D};\theta) = \sum_{i=1}^{N} m_i \log\left[1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}\right] - \sum_{i=1}^{N} (M - m_i)\left(\frac{t_i}{\eta}\right)^\beta .$$

We have

$$\begin{aligned}
\frac{\partial \log L(\mathcal{D};\theta)}{\partial \eta} &= \sum_{i=1}^{N} \frac{m_i}{1 - R(t_i/\theta)} \frac{\partial\left(1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}\right)}{\partial \eta} + \frac{\beta}{\eta}\sum_{i=1}^{N}(M - m_i)\left(\frac{t_i}{\eta}\right)^\beta \\
&= -\sum_{i=1}^{N} \frac{m_i R(t_i/\theta)}{1 - R(t_i/\theta)} \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^\beta + \frac{\beta}{\eta}\sum_{i=1}^{N}(M - m_i)\left(\frac{t_i}{\eta}\right)^\beta .
\end{aligned}$$

This leads to the likelihood equation

$$\sum_{i=1}^{N} t_i^\beta \left[(m_i - M) - m_i \frac{R(t_i/\theta)}{1 - R(t_i/\theta)}\right] = 0 \tag{6}$$

On the other hand

$$\begin{aligned}
\frac{\partial \log L(\mathcal{D};\theta)}{\partial \beta} &= \sum_{i=1}^{N} \frac{m_i}{1 - R(t_i/\theta)} \frac{\partial\left(1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}\right)}{\partial \beta} - \sum_{i=1}^{N}(M - m_i)\left(\frac{t_i}{\eta}\right)^\beta \log\left(\frac{t_i}{\eta}\right) \\
&= \sum_{i=1}^{N} \frac{m_i R(t_i/\theta)}{1 - R(t_i/\theta)} \frac{\partial}{\partial \beta}\left(\frac{t_i}{\eta}\right)^\beta - \sum_{i=1}^{N}(M - m_i)\left(\frac{t_i}{\eta}\right)^\beta \log\left(\frac{t_i}{\eta}\right) \\
&= \sum_{i=1}^{N} \frac{m_i R(t_i/\theta)}{1 - R(t_i/\theta)} \left(\frac{t_i}{\eta}\right)^\beta \log\left(\frac{t_i}{\eta}\right) - \sum_{i=1}^{N}(M - m_i)\left(\frac{t_i}{\eta}\right)^\beta \log\left(\frac{t_i}{\eta}\right)
\end{aligned}$$

Using equation 6, it leads to the likelihood equation

$$\sum_{i=1}^{N} t_i^\beta \log(t_i)\left[(m_i - M) - m_i \frac{R(t_i/\theta)}{1 - R(t_i/\theta)}\right] = 0. \tag{7}$$

Appendix B: EM Algorithm for doubly censored data

We denote $t_{ij} = t_i + \delta_{ij}$ where

- $\mathcal{D} = (t_{ij})$ for $i = 1, ..., N$ and $j = 1, ..., M$ denote complete data,
- $\mathcal{E} = |\delta_{ij}|$ for $i = 1, ..., N$ and $j = 1, ..., M$ denote missing data,
- $\mathcal{S} = (t_i, \mathrm{sgn}(\delta_{ij}))$ for $i = 1, ..., N$ and $j = 1, ..., M$ denote observed data.

The likelihood function for the complete data is

$$L(\mathcal{D}, \theta) = \prod_{i=1}^{N} \left[\prod_{j=1}^{M} f(t_{ij}) \right],$$

with

$$f(t/\theta) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta - 1} \exp \left(- \left(\frac{t}{\eta} \right)^{\beta} \right) \quad \text{for} \quad t \geq 0,$$

and $\theta = (\beta, \eta)$.

The EM algorithm consists of alternating the E step and the M step.

Step E: Denoting θ^- the current vector estimate, this step consist of determining the function

$$Q(\theta/\theta^-) = E_{\theta^-} \left(\log L(\mathcal{D}, \theta)/\mathcal{S} \right),$$

where

$$\log L(\mathcal{D}, \theta) = \sum_{i=1}^{N} \sum_{j=1}^{M} \left[\log \left(\frac{\beta}{\eta} \right) + (\beta - 1) \log \left(\frac{t_{ij}}{\eta} \right) - \left(\frac{t_{ij}}{\eta} \right)^{\beta} \right].$$

We have

$$\begin{aligned}
Q(\theta/\theta^-) &= MN \left[\log(\beta) - \beta \log(\eta) \right] + \sum_{i=1}^{N} \sum_{j=1}^{M} \left[(\beta - 1) E_{\theta^-} \left(\log(t_{ij})/\mathcal{S} \right) - \frac{1}{\eta^{\beta}} E_{\theta^-} \left(t_{ij}^{\beta}/\mathcal{S} \right) \right] \\
&= MN \left[\log(\beta) - \beta \log(\eta) \right] \\
&+ (\beta - 1) \sum_{i=1}^{N} \left[(M - m_i) E_{\theta^-} (\log(t_{ij})/t_{ij} > t_i) + m_i E_{\theta^-} (\log(t_{ij})/t_{ij} < t_i) \right] \\
&- \frac{1}{\eta^{\beta}} \sum_{i=1}^{N} \left[(M - m_i) E_{\theta^-} (t_{ij}^{\beta}/t_{ij} > t_i) + m_i E_{\theta^-} (t_{ij}^{\beta}/t_{ij} < t_i) \right]
\end{aligned}$$

where

$$
\begin{aligned}
E_{\theta^-}(t_{ij}^{\beta}/t_{ij} > t_i) &= \frac{\displaystyle\int_{t_i}^{+\infty} t^{\beta} f(t/\theta^-)dt}{R(t_i/\theta^-)}, \\
E_{\theta^-}(t_{ij}^{\beta}/t_{ij} < t_i) &= \frac{\displaystyle\int_{0}^{t_i} t^{\beta} f(t/\theta^-)dt}{1-R(t_i/\theta^-)}, \\
E_{\theta^-}(\log(t_{ij})/t_{ij} > t_i) &= \frac{\displaystyle\int_{t_i}^{+\infty} \log(t) f(t/\theta^-)dt}{R(t_i/\theta^-)}, \\
E_{\theta^-}(\log(t_{ij})/t_{ij} < t_i) &= \frac{\displaystyle\int_{0}^{t_i} \log(t) f(t/\theta^-)dt}{1-R(t_i/\theta^-)},
\end{aligned}
$$

and

$$
R(t_i/\theta) = \exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right).
$$

Step M: Determine θ^+ maximizing $Q(\theta/\theta^-)$. θ^+ is solution of equation

$$
\frac{\partial Q(\theta^+/\theta^-)}{\partial \theta^+} = 0.
$$

$$
\begin{aligned}
\frac{\partial Q(\theta^+/\theta^-)}{\partial \eta^+} &= -\frac{MN\beta^+}{\eta^+} \\
&+ \frac{\beta^+}{\eta^{+\beta^+ +1}} \sum_{i=1}^{N} \left[(M-m_i)E_{\theta^-}(t_{ij}^{\beta^+}/t_{ij} > t_i) + m_i E_{\theta^-}(t_{ij}^{\beta^+}/t_{ij} < t_i)\right]
\end{aligned}
$$

It leads to the equation

$$
(\eta^+)^{\beta^+} = \frac{1}{MN} \sum_{i=1}^{N} \left[(M-m_i)E_{\theta^-}(t_{ij}^{\beta^+}/t_{ij} > t_i) + m_i E_{\theta^-}(t_{ij}^{\beta^+}/t_{ij} < t_i)\right]. \qquad (8)
$$

$$
\begin{aligned}
\frac{\partial Q(\theta^+/\theta^-)}{\partial \beta^+} &= \frac{MN}{\beta^+} - MN\log(\eta^+) + \sum_{i=1}^{N}\sum_{j=1}^{M} E_{\theta^-}[\log(t_{ij})/\mathcal{S}] \\
&+ \frac{\log(\eta^+)}{(\eta^+)^{\beta^+}} \sum_{i=1}^{N}\sum_{j=1}^{M} E_{\theta^-}\left(t_{ij}^{\beta^+}/\mathcal{S}\right) - \frac{1}{(\eta^+)^{\beta^+}} \sum_{i=1}^{N}\sum_{j=1}^{M} \frac{\partial E_{\theta^-}\left(t_{ij}^{\beta^+}/\mathcal{S}\right)}{\partial \beta^+}
\end{aligned}
$$

It leads to the equation

$$\frac{1}{\beta^+} = -\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M} E_{\theta^-}\left[\log(t_{ij})/\mathcal{S}\right] + \frac{1}{MN(\eta^+)^{\beta^+}}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{\partial E_{\theta^-}(t_{ij}^{\beta^+}/\mathcal{S})}{\partial\beta^+}. \quad (9)$$

Proposition: Solutions of ml equations (2) and (3) are fixed points of the EM algorithm.

Proof: From (8), fixed point of EM verify

$$\eta^\beta = \frac{1}{MN}\sum_{i=1}^{N}\left[(M-m_i)E_\theta\left(t_{ij}^\beta/t_{ij} > t_i\right) + m_i E_\theta\left(t_{ij}^\beta/t_{ij} < t_i\right)\right]$$

and we have from direct calculations

$$E_\theta\left(t_{ij}^\beta/t_{ij} > t_i\right) = \eta^\beta + \left(\frac{t_i}{\eta}\right)^\beta,$$

$$E_\theta\left(t_{ij}^\beta/t_{ij} < t_i\right) = \eta^\beta - \left(\frac{t_i}{\eta}\right)^\beta \frac{R(t_i/\theta)}{1-R(t_i/\theta)},$$

from where

$$\sum_{i=1}^{N} t_i^\beta\left[(M-m_i) - m_i\frac{R(t_i/\theta)}{1-R(t_i/\theta)}\right] = 0,$$

which is equation (3.2).

¿From (9), fixed point of EM verify

$$\begin{aligned}\frac{1}{\beta} &= -\frac{1}{MN}\sum_{i=1}^{N}\left[(M-m_i)E_\theta\left(\log(t_{ij})/t_{ij} > t_i\right) + m_i E_\theta\left(\log(t_{ij})/t_{ij} < t_i\right)\right]\\ &+ \frac{1}{MN\eta^\beta}\sum_{i=1}^{N}\left[(M-m_i)E_\theta\left(t_{ij}^\beta\log(t_{ij})/t_{ij} > t_i\right) + m_i E_\theta\left(t_{ij}^\beta\log(t_{ij})/t_{ij} < t_i\right)\right],\end{aligned}$$

and we have, letting $u = (t/\eta)^\beta$ and using integration by parts.

$$\begin{aligned}E_\theta(\log(t_{ij})/t_{ij} > t_i) &= \frac{1}{R(t_i/\theta)}\int_{t_i}^{+\infty}\log(t)\left(\frac{\beta}{\eta}\right)\left(\frac{t}{\eta}\right)^{\beta-1}\exp\left(-\left(\frac{t_{ij}}{\eta}\right)^\beta\right)dt\\ &= \frac{1}{R(t_i/\theta)}\int_{(\frac{t_i}{\eta})^\beta}^{+\infty}\log\left(\eta u^{\frac{1}{\beta}}\right)\exp(-u)du\\ &= \log(\eta) + \frac{1}{\beta R(t_i/\theta)}\int_{(\frac{t_i}{\eta})^\beta}^{+\infty}\log(u)\exp(-u)du,\end{aligned}$$

$$
\begin{aligned}
E_\theta(\log(t_{ij})/t_{ij} < t_i) &= \frac{1}{1-R(t_i/\theta)} \int_0^{t_i} \log(t) \left(\frac{\beta}{\eta}\right) \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t_{ij}}{\eta}\right)^{\beta}\right) dt \\
&= \frac{1}{1-R(t_i/\theta)} \int_0^{(\frac{t_i}{\eta})^\beta} \log\left(\eta u^{\frac{1}{\beta}}\right) \exp(-u) du \\
&= \log(\eta) + \frac{1}{\beta(1-R(t_i/\theta))} \int_0^{(\frac{t_i}{\eta})^\beta} \log(u) \exp(-u) du, \\
E_\theta(t_{ij}^\beta \log(t_{ij})/t_{ij} > t_i) &= \frac{1}{R(t_i/\theta)} \int_{t_i}^{+\infty} t^\beta \log(t) \left(\frac{\beta}{\eta}\right) \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t_{ij}}{\eta}\right)^{\beta}\right) dt \\
&= \frac{\eta^\beta}{R(t_i/\theta)} \int_{(\frac{t_i}{\eta})^\beta}^{+\infty} u \log\left(\eta u^{\frac{1}{\beta}}\right) \exp(-u) du \\
&= \frac{\eta^\beta \log(\eta)}{R(t_i/\theta)} \int_{(\frac{t_i}{\eta})^\beta}^{+\infty} u \exp(-u) du \\
&+ \frac{\eta^\beta}{\beta R(t_i/\theta)} \int_{(\frac{t_i}{\eta})^\beta}^{+\infty} u \log(u) \exp(-u) du \\
&= \eta^\beta \log(\eta) \\
&+ \frac{\eta^\beta}{\beta} + \eta^\beta \left(\frac{t_i}{\eta}\right)^\beta \log(t_i) + \frac{\eta^\beta}{\beta R(t_i/\theta)} \int_{(\frac{t_i}{\eta})^\beta}^{+\infty} \log(u) \exp(-u) du, \\
E_\theta(t_{ij}^\beta \log(t_{ij})/t_{ij} < t_i) &= \frac{1}{1-R(t_i/\theta)} \int_0^{t_i} t^\beta \log(t) \left(\frac{\beta}{\eta}\right) \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t_{ij}}{\eta}\right)^{\beta}\right) dt
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\eta^\beta}{1-R(t_i/\theta)} \int_0^{(\frac{t_i}{\eta})^\beta} u \log\left(\eta u^{\frac{1}{\beta}}\right) \exp(-u) du \\
&= \frac{\eta^\beta}{1-R(t_i/\theta)} \left[\log(\eta) \int_0^{(\frac{t_i}{\eta})^\beta} u \exp(-u) du + \frac{1}{\beta} \int_0^{(\frac{t_i}{\eta})^\beta} u \log(u) \exp(-u) du\right] \\
&= \eta^\beta \log(\eta) + \frac{\eta^\beta}{\beta} - \eta^\beta \left(\frac{t_i}{\eta}\right)^\beta \log(t_i) \frac{R(t_i/\theta)}{1-R(t_i/\theta)} \\
&+ \frac{\eta^\beta}{\beta(1-R(t_i/\theta))} \int_0^{(\frac{t_i}{\eta})^\beta} \log(u) \exp(-u) du.
\end{aligned}
$$

Thus, gathering together all those equations, we get

$$
\frac{MN}{\beta} = -\sum_{i=1}^{N} (M-m_i) \left[\log(\eta) + \frac{1}{\beta R(t_i/\theta)} \int_{(\frac{t_i}{\eta})^\beta}^{+\infty} \log(u) \exp(-u) du\right]
$$

$$
\begin{aligned}
&- \sum_{i=1}^{N} m_i \left[\log(\eta) + \frac{1}{\beta(1-R(t_i/\theta))} \int_0^{\left(\frac{t_i}{\eta}\right)^\beta} \log(u)\exp(-u)du \right] \\
&+ \sum_{i=1}^{N} (M-m_i) \left[\log(\eta) + \frac{1}{\beta} + \left(\frac{t_i}{\eta}\right)^\beta \log(t_i) + \frac{1}{\beta R(t_i/\theta)} \int_{\left(\frac{t_i}{\eta}\right)^\beta}^{+\infty} \log(u)\exp(-u)du \right] \\
&+ \sum_{i=1}^{N} m_i \left[\log(\eta) + \frac{1}{\beta} - \left(\frac{t_i}{\eta}\right)^\beta \log(t_i) \frac{R(t_i/\theta)}{1-R(t_i/\theta)} + \frac{1}{\beta(1-R(t_i/\theta))} \int_0^{\left(\frac{t_i}{\eta}\right)^\beta} \log(u)\exp(-u)du \right],
\end{aligned}
$$

and finally, we get the equation

$$
\sum_{i=1}^{N} t_i^\beta \log(t_i) \left[(m_i - M) - m_i \frac{R(t_i/\theta)}{1-R(t_i/\theta)} \right] \quad = \quad 0
$$

which is (3.3).

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399